DUPLICATE

METEOROLOGICAL OFFICE

Forecasting Research

Forecasting Research Division
Technical Report No. 192

Report on experiments to improve precipitation nowcasts in Nimrod

by

W H Hand and D Steadman

April 1996

Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

ORGS UKMO F

National Meteorological Library
FitzRoy Road, Exeter, Devon. EX1 3PB

Forecasting Research
Technical Report No. 192

Report on experiments to improve precipitation nowcasts in Nimrod

by

W H Hand and D Steadman

April 1996

Forecasting Research
Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

© Crown Copyright 1996

This paper has not been peer-reviewed and should be regarded as an Internal Report from the Meteorological Office. Permission to quote from it should be obtained from the above Meteorological Office division.

Report on experiments to improve precipitation nowcasts in Nimrod.

by

W.H.Hand and D.Steadman

0. Summary

Experiments have been performed to improve the extrapolation and model merging components of the rainfall forecast in Nimrod. The experiments were conducted in semi-real-time on the Nimrod development workstation to produce a comprehensive set of verification data and statistics and to enable a fair comparison with operational results. For the extrapolation, tests were made of extending the object motion analysis to the entire Nimrod area. This gave results which were on average no worse than the operational methods and on some occasions were better. As well as extending the area over which objects are analysed, experiments were performed in using object envelopes to determine movement of rainfall regions over timescales longer than an hour. The conclusion was that extrapolation using object envelopes matched 3 hours apart was beneficial in some situations. Subjective assessment revealed these to be mainly showery types where the areas affected by showers remained constant. Matching envelopes more than 3 hours apart proved to be detrimental, producing forecasts which were, on average, much too slow. A new method of extrapolating heavy rain has also been devised and tested. The method analyses the recent movement of heavy rain as well as the overall rain area and produces separate forecasts of heavy and light rain. These forecasts are combined in an optimum way which has been tuned by experiment. The combined (merged) extrapolation forecasts produced results significantly better than the operational system for both heavy and light rain.

A new method of merging model rainfall forecasts with extrapolation forecasts has been devised. The method seeks to choose the best extrapolation forecast, operational (extended area), envelope or combined (merged) and merge this with the latest available UM rainfall forecast. The merging is done using weighted combinations of the UM and extrapolation forecast with the weights determined by the recent performance of each method according to lead time. The new method improves upon the operational scheme and produces a smoother forecast sequence. However, subjective assessment has revealed a tendency for more light rain early in the forecast, particularly when the extrapolation has been performing poorly.

Recommendations have been made that the method of combining heavy and light rain extrapolation forecasts be implemented into Nimrod as soon as possible.This change is to be combined with increasing the region over which object motion is analysed to the whole of the Nimrod domain. Although the new UM merging method results are encouraging, further tuning or modifications are necessary to improve the shorter range nowcasts. However, implementation of an improved model merging scheme into Nimrod is still envisaged.

1. Introduction

Nimrod precipitation rate nowcasts are generated by merging the most recent total precipitation rate forecast from the mesoscale version of the Unified Model (UM), with an extrapolation forecast. The extrapolation is made by analysing recent movement of large rainfall objects inside the UK radar area, (Conway et al,1993). The motion vectors are derived by advecting objects analysed one hour before to the current time using a range of model wind fields at different levels and times and single vectors. The model wind level or single vector which gives the best match at T+0 using cross-correlation is then chosen to make the forecast. Small rainfall objects (< 150 radar pixels) and rainfall objects analysed outside the radar area are assigned the same motion method as the nearest large object. Rainfall rate extrapolations are merged with the UM by adding a weighted fraction of the extrapolation forecast with a weighted fraction of the UM forecast. The weights vary according to forecast lead time, the accuracy of the UM forecast verifying at T+0, and the expected accuracy of the extrapolation using cross-correlation as the performance measure. Before merging, the rainfall rate is transformed to a log variable so that differences in rainfall rate between, say, 8 and 16 mm/h become equally as important as those between 1 and 2 mm/h. The merging weights are calculated such that increasing weight is given to the UM with increasing lead time and corresponding lower weight to the extrapolation forecast. This means that by T+6 the forecast predominantly reflects the UM precipitation field on a 5km grid.

Operationally, only objects within the useable radar area are used to determine motion vectors. Although at first sight this appears to be restrictive, it must be remembered that if model winds are chosen to advect an object then objects outside the radar boundary will still be advected with the UM winds at the object location. However, if linear extrapolation is chosen then the same vector may not be relevant for distant objects. By extending the object motion analysis to cover the entire Nimrod region it was expected that more accurate motion vectors could be derived for objects outside the radar area. Also very large, but single objects where the radar only samples part of them, were expected to be treated more accurately. It is also desirable to extend the area of the object analysis to reap the full benefit of improvements in techniques for analysing rainfall outside the radar area.

Since Nimrod has been running operationally it has occasionally been noted, particularly in some showery situations, that the rainfall forecast has been too fast. On these occasions, significant decay occured in reality on the leading edge of the rainfall region as the forecast progressed, thus maintaining the main rain region in approximately the same place. The reason for the forecast errors in these situations is that the analysed motion vectors relate to a one hour of movement of precipitation, which may not be appropriate for a longer time-scale forecast where decay takes place after an hour. A good example of this is the restricted inland penetration of showers in winter. Matching entire rainfall regions (envelopes) which enclose non-precipitating pixels as opposed to just the precipitating objects should produce improvements in these situations since previous work in this area suggested that envelope velocities are generally less than object velocities. However, further experiments were required to ascertain the general applicability of using envelope velocities.

In the operational scheme for determining the motion of rainfall objects, variation in rainfall rate intensity is not used to distinguish objects. However, it is well known that moderate or heavy precipitation often moves in a different direction to the general rain area. For example, frontal rainbands often move along fronts, whereas the main rain area remains tied to and moves with the front. Deep convection occuring in a directionally-sheared wind environment usually means that the long term storm movement may be substantially different from that of the short-lived convective cells which can give very heavy downpours. The aim of the experiments was to distinguish between heavy rain and light rain in determining object motion and to quantify any subsequent improvement in the prediction of heavy rain.

The present method of merging the extrapolation forecast with the UM is not optimal in the sense that it is assumed that the model is always likely to be better in the latter stages of a 6 hour forecast. However,

the information content of advection forecasts can be highly variable, often depending on the synoptic situation. An experiment has been carried out to merge the advection and UM forecasts in a way that is more directly related to their relative recent performance. Also, in assessing performance, the accuracy of the rainfall rates as well as rainfall location has been taken into account.

2. Extrapolation experiments

All of the experiments described in this section were conducted in "near real-time" using operational software on the Nimrod development machine NIMDEV. The reasons for doing this as opposed to performing casestudies were to expose the new techniques to a large variety of conditions and to ensure that the same UM datafiles were used by the experiment programs as were available to the operational ones. The experiments described in sections 2.2, 2.3, and 2.4 were run on data in the period from 0600 UTC 22/12/1995 to 0545 UTC 16/01/96. In this period a wide variety of rainfall events occured in a range of synoptic conditions. Forecasts were run from hourly data with object motion analyses every 15 minutes. It was not possible to reproduce the half-hourly cycle that is run on Nimrod as it was necessary to run and verify the experiments concurrently and the NIMDEV HP workstation was not fast enough to process the additional runs in real time. All of the experimental forecasts were compared objectively with radar analyses using the same programs used operationally which provide a comprehensive range of skill scores. Statistics covering the entire period have been produced and these are presented in this paper. As well as the objective verification, all forecasts verifying at 1200 UTC each day were saved for subjective assessment.

2.1 Objective verification

Three skill scores were chosen to judge the success of the experiments; Hit Rate (HR), False Alarm Rate (FAR), and Root Mean Square Factor Or (RMSF_OR). HR and FAR provide measures of rainfall location skill and RMSF_OR provides a measure of rainfall rate skill combined with location. HR is defined as the number of occasions an event has been observed and forecast divided by the number of occasions it has been observed. An event in this case is a rain rate above a specified threshold and an occasion is a pixel occurence of an event. FAR is defined as the number of occasions an event has been forecast and not observed divided by the number of occasions it has been forecast. HR and FAR are expressed as percentages and the ideal score for HR is 100 and for FAR is 0. In order to define RMSF_OR we must first define RMSF.

Given forecast and analysed events F and G then,

$$RMSF = \log^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\log\frac{F_i}{G_i}\right)^2\right)^{1/2} \qquad (F,G > 0)$$

In this definition all pixels where rain has been forecast but not observed (F>0, G=0) and where rain has been observed but not forecast (F=0, G>0) are ignored. The RMSF_OR statistic takes these into account by setting F or G to half the threshold value chosen for an event. This means that in those regions outside the intersection of F and G, errors will occur depending on the size of the under or over-prediction. The ideal score for RMSF_OR is 1.0.

HR and FAR are calculated at the full Nimrod resolution of 5km. RMSF_OR is calculated at 15km resolution using 3x3 Nimrod pixel averaging. This is approximately the same resolution as the UM. These methods were chosen to satisfy the Nimrod project User Requirement.

2.2 Experiment to extend region of object motion analysis to whole of Nimrod area.

To perform this experiment, rainfall objects were analysed over the whole of the Nimrod domain. Although the size of a large object was retained at 150 pixels the number of large objects allowed per analysis was increased from 5 to 7, which was the maximum permitted by the memory size of NIMDEV.

Summary statistics for the experiment period are presented in table 1 below.

Table 1). HR, FAR and RMSF_OR for the operational (O) and extended area (A) extrapolation forecasts over the test period for three threshold rainfall rates. HR and FAR are expressed as percentages (* = 100) and RMSF_OR has been multiplied by 10.

Lead times (min)	Threshold = 0.125 mm/h						Threshold = 2.0 mm/h						Threshold = 10.0 mm/h					
	HR		FAR		RMSF OR*10		HR		FAR		RMSF OR*10		HR		FAR		RMSF OR*10	
	O	A	O	A	O	A	O	A	O	A	O	A	O	A	O	A	O	A
T+360	29	31	64	64	74	75	10	10	91	91	29	29	0	0	*	*	25	25
T+300	34	36	60	59	74	74	11	11	90	90	29	29	0	0	*	*	25	25
T+240	41	42	55	54	71	71	13	14	88	87	29	29	0	0	*	*	25	25
T+180	50	50	48	47	67	67	18	18	84	84	28	28	0	0	*	*	25	25
T+120	60	60	39	39	60	60	26	26	77	77	27	27	0	0	*	*	25	25
T+90	66	66	33	33	54	54	32	33	72	71	26	26	4	5	95	94	24	24
T+60	74	74	26	26	46	46	42	43	62	62	24	24	6	6	92	91	24	24
T+30	85	85	17	16	34	33	61	61	45	45	21	21	18	18	78	76	22	22

It is apparent from table 1 that there is hardly any difference in forecast performance between the operational runs and those forecasts from the extended object motion analyses. At only one lead time (T+360 for RMSF at 0.125 mm/h) was the new scheme worse than the operational system for any skill measure. Subjectively, it was also difficult to find major differences between the operational forecasts and the experimental ones, particularly before T+3. However, figure 1 does show a clear improvement in the experimental extrapolation forecast for T+3 over France. This can be attributed to an improved analysis of the motion of this rainfall area in the test scheme outside the radar area. Figure 1 will be referred to again later.

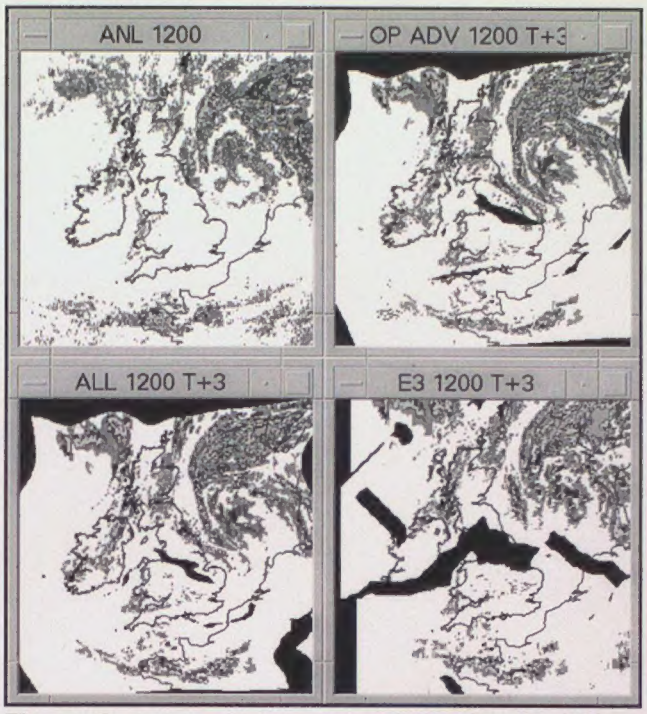

Figure 1. Rainfall images for 1200 UTC 25/12/95. "ANL" is the Nimrod operational analysis, "OP" is the Nimrod operational 3 hour extrapolation forecast for 1200 UTC. "ALL" is the forecast made by analysing objects over the entire Nimrod area and "E3" is a forecast using 3 hour envelope velocities (see text for details). The black regions are areas of missing data.

2.3 Experiment in using envelope velocities for extrapolation

To perform this experiment large object envelopes were matched as opposed to the rainfall objects. Envelopes were analysed over the entire Nimrod region. The envelope is defined as those pixels contained within an region enclosed by a line joining all the edge pixels. Thus for an area of showers, the envelope represents the broad region affected by showers as opposed to individual shower clusters.

In matching envelopes and calculating envelope velocities we are trying to determine the movement of large precipitation regions within which there will be a smaller scale distribution of rainfall varying on a shorter timescale than the envelope. In the case of showery regions, the 1 hour time lag used operationally to determine object motion vectors is a compromise between picking up the movement of individual showers and the general showery area.

In the experiment, envelope velocities were calculated using time lags of 3 hours and 6 hours and forecasts made and compared. The objective scores for these experiments are shown in table 2 below and are compared with those from the operational system using the full Nimrod area.

Table 2). HR, FAR and RMSF_OR for the full area (A) and envelope forecasts using a 3 hour correlation time lag (E3) and a 6 hour lag (E6) over the test period for a threshold rainfall rate of 0.125 mm/h. HR and FAR are expressed as percentages.

Thresh 0.125 mm/h	HR			FAR			RMSF_OR		
	A	E3	E6	A	E3	E6	A	E3	E6
T+360	31	28	25	64	64	66	7.50	7.25	7.34
T+300	36	33	30	59	60	62	7.39	7.27	7.30
T+240	42	39	36	54	55	58	7.15	7.07	7.28
T+180	50	47	44	47	49	52	6.69	6.72	6.86
T+120	60	56	53	39	41	43	5.99	6.21	6.54
T+90	66	63	60	33	35	37	5.42	5.78	6.12
T+60	74	71	69	26	28	30	4.62	5.00	5.31
T+45	79	76	75	22	24	25	4.08	4.45	4.73
T+30	85	82	82	16	18	19	3.33	3.69	3.91
T+15	91	90	89	11	12	12	2.40	2.64	2.76

For HR and FAR the operational forecast over the extended area produces the best results. Of the envelopes, the one using a 3 hourly correlation time lag is superior. For the RMSF_OR statistic The E3 forecasts outperformed the A forecasts after T+180 and the E6 forecasts after T+240. E3 forecasts were always better than E6 forecasts. The lower scores after T+180 for the E3 forecasts appears to contradict the results using HR and FAR. However, beyond T+180 the difference in FARs between A and E3 is very small. This suggests that the envelope forecast has less tendency to overpredict rain at longer ranges than the operational forecast. Subjective examination confirmed that envelope forecasts are generally slower and that this was beneficial in a few cases. An example of this is shown in figure 1. The forecast of the showery regions over Ireland and the Highlands is better in the envelope forecast than the operational. However, the operational forecast of the movement of the vortex in the North Sea is faster and more correct. Up to T+180 envelope scores were consistently worse than the operational ones, which implies that the value of envelope techniques is not realised until later in the forecast period, when extrapolation of large features becomes important. However, envelope forecasts will only be useful in certain situations and would be detrimental to Nimrod if they replaced the operational methods particularly in the early stages of a forecast when extrapolation has a higher weight than the UM. This implies that if they are to be used then some objective method of determining when they should be applied needs to be found. An experiment on these lines is presented later.

2.4 Experiments in using different extrapolation velocities according to rainfall rate

Rainfall rates greater than 1mm/h usually occur as a result of the release of instabilities in the atmosphere which vary on very short space and timescales. Typical examples are showers and frontal rainbands. If we are to achieve more accurate very short range forecasts of heavy rain then we must at least be able to predict the expected movement of the heavy rain areas. Because of the high temporal variation a correlation time lag of one hour is too large to derive the motion vectors of heavy rain pixels. Therefore,

the experiment was set up to use a time lag of 15 minutes, which is the minimum time between successive Nimrod analyses. Several intensity thresholds for heavy rain were tried, however, 0.5 mm/h proved to be optimal as higher thresholds tended to produce more and smaller objects making spurious correlations more likely. However, using a threshold of 0.5 did require that the minimum size for a large object be set to 50 pixels. To perform the experiment the object motion analysis was run twice, thresholding the rainfall rate at the operational value of 1/32nd mm/h with a correlation time lag of 60 minutes, and again at the higher threshold of 0.5 mm/hr with a time lag of 15 minutes. For the higher threshold correlation only linear extrapolation vectors were tested to derive motion vectors, since over a 15 minute period these were found to give the best estimate of the movement of heavy rain areas. Two forecasts were run from each analysis, a standard forecast (S) and a forecast with rain thresholded at 0.5 mm/h (HI). Given that high intensity rainfall objects were going to move with different velocities from the light rain, a method of combining (merging) the forecasts was devised as follows:

1. All merged forecast pixels were initialized to zero.

2. The average rain rate of all pixels in the analysis below the high threshold (0.5 mm/h) was calculated.

3. Where the HI forecast had pixels higher than the threshold, (some were not because of smudging in the advection scheme), the two forecasts were merged using a weighted combination. The weights varied linearly with forecast lead time such that full weight was given to the HI forecast at T+15 and zero weight at 15 minutes past the end of the forecast period of the HI forecast.

4. At pixels containing heavy rain from the S forecast but not in the HI forecast, a weighted combination of the average rate and the standard rate was set, merging towards the S forecast rate at the end of the forecast period using the same linear weights as before.

5. All light rain pixels present in the S forecast were added to the merged forecast if the corresponding merged forecast pixels were dry.

6. The remaining dry merged forecast pixels were then checked and if either the corresponding pixel in the S or HI forecast contained a missing data value then the merged forecast pixel was set to contain a missing data value.

The operation and effect of this procedure is best illustrated by two examples:

(a) Rainband within a front

Rainbands can usually be found in all active fronts and are a consequence of the release of atmospheric instability by a variety of mechanisms. Many rainbands occur as a result of instabilities at medium levels. In a frontal situation, the rain area will move with the front, usually with a velocity close to the geostrophic wind near the surface. However, rainbands are often observed to move along the front with a velocity corresponding to winds in the warm air at medium or sometimes fairly high levels, which are substantially veered from the winds lower down. Nimrod in its present formulation would move all the rain together, preserving the relative position of the heavy rain to the front. The effect of the experimental scheme over a period of an hour would be as shown schematically in figure 2.

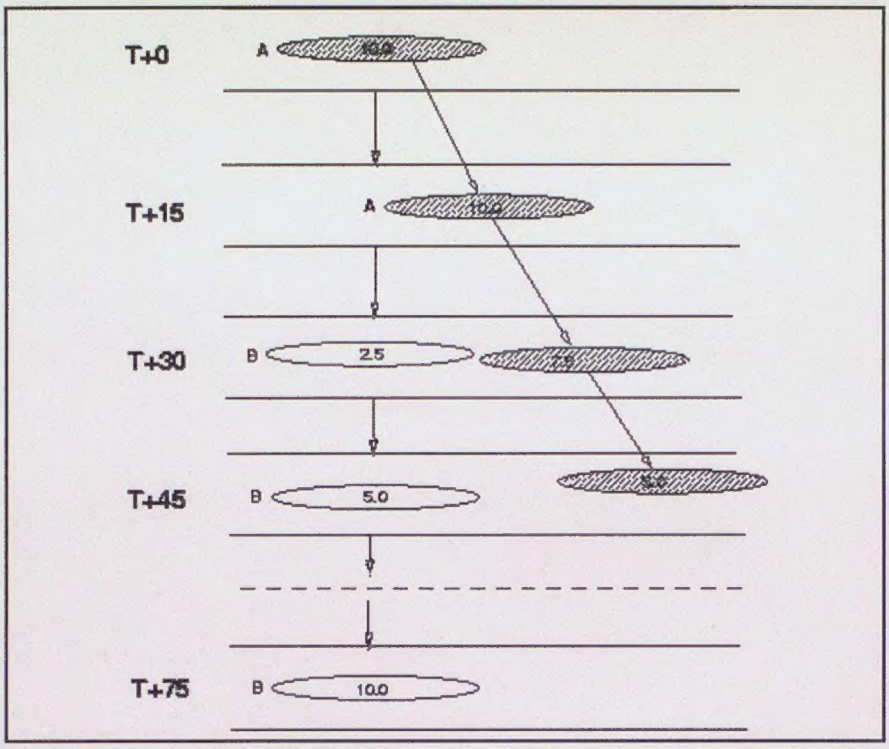

Figure 2. Schematic showing a region of frontal rain at different lead times. The shaded region A shows the movement of a rainband in the HI forecast and the rain rate after merging. The non-shaded region B is a region of rain where the S forecast would have had heavy rain.

Figure 2 shows the progression of a frontal rain area with a rainband (area of rain = 10 mm/h) shaded. (The rest of the rain is assumed to be very light, and the average has not been included for simplicity). The arrows showing the movement of the shaded area (A) indicates the real movement of the rainband relative to the ground. The other arrows indicate the movement of the front relative to the ground. If we assume that the two runs of the object motion analysis derive the correct motion vectors for the frontal rain and the rainband then by T+15 the rainband has moved to the right as the frontal rain has moved straight on. Since we are merging over a period of an hour the weights for the heavy rain would be 1.0 at T+15, 0.75 at T+30, 0.5 at T+45, 0.25 at T+60 and 0.0 at T+75. At T+15 the heavy rain in the S forecast is not used, so we have all the rainband from the HI forecast. At T+30 the original rainband has been advected straight on with the front in the S forecast, and is represented by area B with a rate of 2.5 mm/hr. The shaded rainband from the HI forecast is reduced to 7.5 mm/hr. This process of increase and reduction continues until T+75 when we have reproduced the standard Nimrod extrapolation forecast prior to merging with the model.

(b) Heavy showers

This is different to the frontal situation in that in showery situations there is usually proportionately more heavy rain in the overall rainy area. Therefore, the interpretation of the effect of the new scheme is going to be different. Figure 3 illustrates the process schematically and again the average value of the light rain has not been included for simplicity.

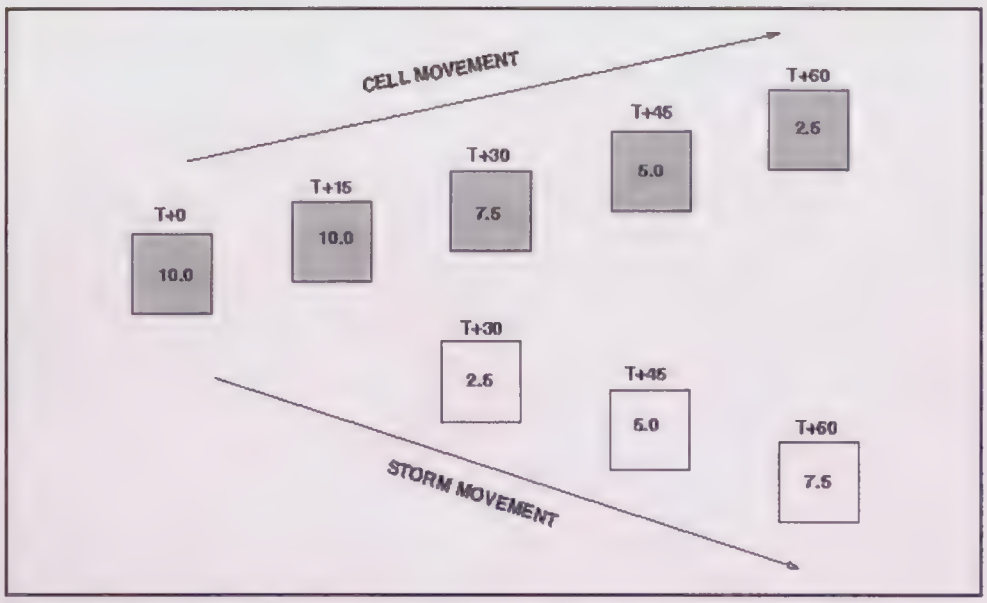

Figure 3. Schematic showing the movement and change of rainfall rate of an area of heavy rain in the S and HI forecasts. The HI forecast is shaded and can be taken to represent cell movement. The S forecast is non-shaded and represents storm movement with a new cell growing in intensity as the rainfall rate increases.

The meaning of the shading is the same as it was in the previous figure, except that the hourly derived vectors from the S forecast will give the storm movement and the HI vectors will give cell movement. This is quite realistic, since the diagram shows the cell decaying and the storm increasing in intensity with time. In reality new cells will be replacing old cells to give the storm movement, the increase in storm intensity is a reasonable simulation of this effect.

In the experiment the HI forecast was merged with the S forecast over four time periods; 1, 2, 3, and 6 hours. This was to determine the optimum lead time or weights for the merging. As well as these four forecasts the HI thresholded forecast without merging was also verified. The results are summarised in tables 3 and 4 where the A scores are shown for comparison.

Table 3). HR, FAR and RMSF_OR for the full area (A), thresholded (HI), and merged A and HI forecasts over 1, 2, 3, and 6 hours (M1, M2, M3, and M6) for a threshold rainfall rate of 0.125 mm/h. HR and FAR are expressed as percentages. The best RMSF_OR scores for each lead time have been underlined.

Lead times (min)	Threshold = 0.125 mm/h																	
	Hit Rate						False Alarm Rate						RMSF_OR x 100					
	A	HI	M1	M2	M3	M6	A	HI	M1	M2	M3	M6	A	HI	M1	M2	M3	M6
T+360	31	22	-	-	-	32	64	62	-	-	-	64	750	755	-	-	-	747
T+300	36	26	-	-	-	44	59	56	-	-	-	60	739	778	-	-	-	689
T+240	42	32	-	-	-	51	54	50	-	-	-	55	715	747	-	-	-	646
T+180	50	38	-	-	55	59	47	42	-	-	48	49	669	723	-	-	645	589
T+120	60	45	-	66	68	68	39	32	-	40	41	41	599	654	-	564	531	518
T+90	66	50	-	73	73	73	31	22	-	36	36	36	542	598	-	486	475	470
T+60	74	56	79	79	79	79	36	26	29	29	29	29	462	507	425	405	402	404
T+45	79	59	83	83	83	83	42	32	25	25	25	25	408	449	365	359	360	363
T+30	85	63	87	87	87	87	16	9	18	18	18	18	333	372	297	300	302	305
T+15	91	63	92	92	92	92	11	4	12	12	12	12	240	284	233	233	233	233

Table 4). HR, FAR and RMSF_OR for the full area (A), thresholded (HI), and merged A and HI forecasts over 1, 2, 3, and 6 hours (M1, M2, M3, and M6) for a threshold rainfall rate of 2.0 mm/h. HR and FAR are expressed as percentages. The best RMSF_OR scores for each lead time have been underlined.

Lead times (min)	Threshold = 2.0 mm/h																	
	Hit Rate						False Alarm Rate						RMSF_OR x 100					
	A	HI	M1	M2	M3	M6	A	HI	M1	M2	M3	M6	A	HI	M1	M2	M3	M6
T+360	10	7	-	-	-	9	91	93	-	-	-	91	291	298	-	-	-	290
T+300	11	9	-	-	-	9	90	90	-	-	-	89	290	296	-	-	-	284
T+240	14	13	-	-	-	12	87	87	-	-	-	85	286	290	-	-	-	278
T+180	18	18	-	-	17	15	84	82	-	-	83	79	280	281	-	-	278	272
T+120	26	25	-	25	22	22	77	74	-	75	71	71	268	269	-	264	259	262
T+90	33	32	-	29	28	30	71	66	-	65	63	64	258	259	-	249	249	254
T+60	43	43	40	39	40	41	62	54	57	52	52	53	241	238	233	231	233	236
T+45	50	50	46	47	48	49	54	46	44	43	44	44	228	226	218	221	222	224
T+30	61	60	58	59	59	60	45	35	33	34	34	35	206	201	197	199	200	201
T+15	74	72	72	72	72	72	31	22	22	22	22	22	171	166	166	166	166	166

In table 3 all of the scores for the HI forecast are worse than those for A. This was because all rain with a rate below 0.5 mm/h was set to zero in the thresholding. However, for HR all of the merged extrapolation forecast scores (M1 - M6) were better than A. There were only small variations between the merged forecasts. For FAR the A scores were slightly better. This was expected since the area of precipitation in the merged forecasts is slightly greater than that in the A forecasts. The difference depending on the divergence of the heavy rain from the lighter rain. However, when we look at rates above 2.0 mm/h in table 4 we see that the HR scores for the merged forecasts were slightly worse than in A, whereas the FAR showed a marked improvement, especially up to T+120. The improvement is greatest in the first hour of the forecast where more weight is given to the HI forecast in the merging for M2, M3 and M6. Towards the end of the merging periods, for example, at T+60 in M1 the scores deteriorate as increasing weight is given to the A forecast. It is clear from these results that the main benefit of the merged extrapolation forecast was in not over-predicting heavy (> 2.0 mm/h) rain. The RMSF_OR scores also justify this conclusion. All of the scores from the merged forecasts in each threshold rate show an improvement over the corresponding A forecast. The best RMSF_OR scores for each lead time have been underlined in both tables. After T+120 the M6 forecast performed the best, but before that it was either M1, M2, or M3 depending on lead time and threshold value used in the verification. It is interesting to compare the best RMSF_OR scores for the merged forecasts using a threshold of 0.125 with the weights given to the HI forecast component. In table 5 below we see that beyond T+15 and up to T+120 a weight for the HI rainfall forecast of 0.75 gives the best result. This result formed the basis for a refinement to the merging procedure which is discussed later. For rates above 2.0 mm/h a weight between 0.5 and 0.75 seems optimal.

Table 5. Weights given to the HI forecast component in each of the merged forecasts corresponding to lead time. Those weights which gave the best RMSF_OR score in the merged forecasts at a threshold of 0.125 mm/h have been underlined.

Weights	M1	M2	M3	M6
T+360	0	0	0	0.04
T+300	0	0	0	0.21
T+240	0	0	0	0.37
T+180	0	0	0.08	0.54
T+120	0	0.13	0.42	0.71
T+90	0	0.37	0.58	0.79
T+60	0.25	0.63	0.75	0.88
T+45	0.50	0.75	0.83	0.92
T+30	0.75	0.87	0.92	0.96
T+15	1.00	1.00	1.00	1.00

For a subjective assessment of the performance of the merged forecasts an example from the trial period is shown in figure 4 below. In this figure are shown three forecasts of rain thresholded at 0.5 mm/h all verifying at 1200 UTC on 30/12/95. The A forecast has a number of significant (for one hour ahead)

errors. The rain in the Irish Sea to the west of Anglesey is too far north. The curved band of rain in Cardigan bay has been advected too fast northwestwards. The band of freezing rain, sleet and snow streching from Hampshire northwest to the Severn Estuary has also been advected too fast north and west. In the HI forecast the rain in the Irish sea and in Cardigan Bay is much better positioned. The line of precipitation from the Severn Estuary to Hampshire has also been better handled, although it is still a little too far north and there is too much precipitation in S. Wales. The effect of the merging procedure can be seen in the M6 forecast where a line of pixels from the A forecast has been added to to the northern edge of the band west of Anglesey in the HI forecast, which was beneficial. Further south, a few pixels have also been added but this time erroneously. This example illustrates the often beneficial effect of the merging process in the forecasts during the trial period and backs up the objective scores.

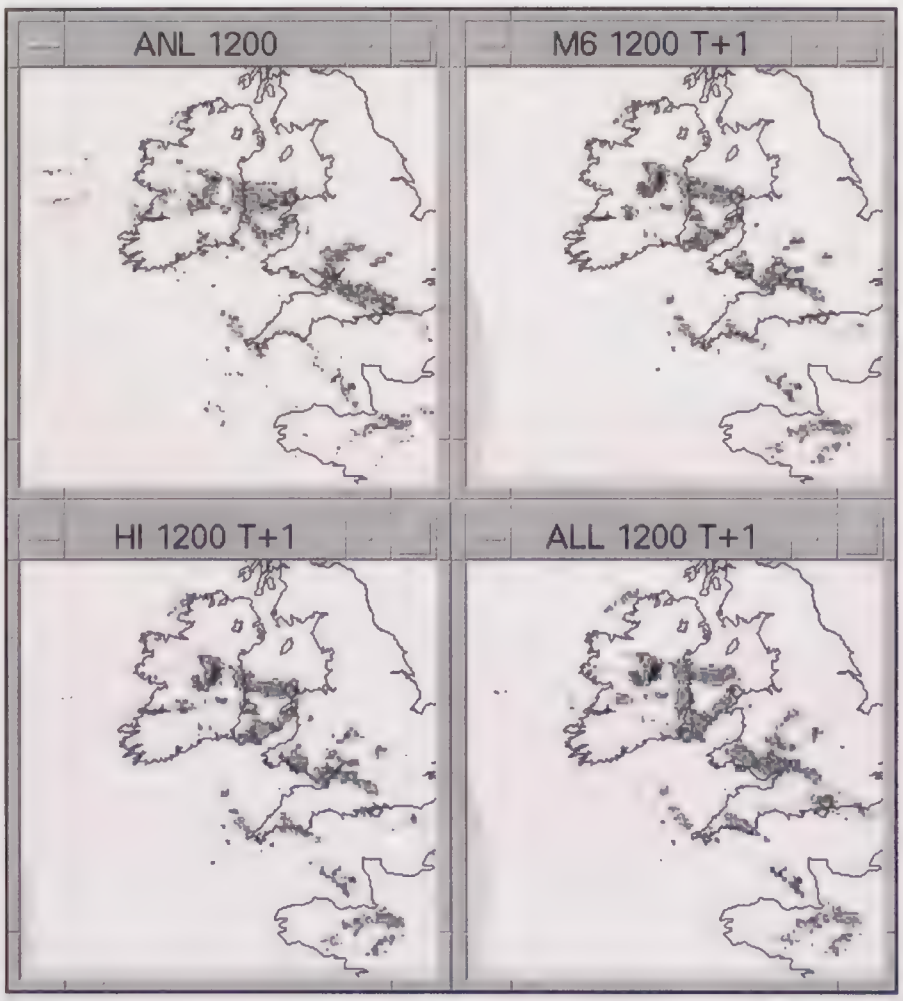

Figure 4. Nimrod analysis (ANL) for 1200 UTC 30/12/95 and 3 T+1 forecasts verifying for the same time. "ALL" is the A forecast, "HI" is the thresholded forecast and "M6" is the merged A and HI forecast. The forecast images show precipitation with rates > 0.5 mm/h.

2.5 Summary of extrapolation experiments.

The main conclusions are:

1. Extending the motion analysis to include rainfall objects outside the useable radar boundary produces forecasts which are on average no worse than the operational ones and on occasions can be expected to be better. (See figure 5a and 5b below).

2. The use of envelope velocities can produce improvements in forecast performance, however, only in a subset of weather types. Envelope velocities derived from matching objects 6 hours apart are clearly worse than those using a time difference of 3 hours (figure 5a and 5b). It was also evident that envelope velocities tended on average to be less detrimental beyond T+240.

3. The method of merging forecasts produced using motion vectors according to rainfall value gave a significant improvement over operational forecasts. Merging over 6 hours (M6) and using HR for rain greater than 0.125 mm/h as the skill measure (figure 5a), gave an improvement of about an hour in extrapolation skill after T+60. However, FARs were slightly worse than the operational runs, except for rain greater than 2.0 mm/h (figure 5d) where all the merged forecasts showed a significant improvement over the operational/ extended area runs. Conversely, at this higher threshold, HRs were slightly worse than the operational/ extended area runs (Table 4). However, RMSF_OR scores for both threshold ranges (tables 3 and 4) showed that all the merged forecasts gave better results than the operational/ extended area runs.

4. Comparison of the individual merged runs (M1, M2, M3 and M6) indicated that optimal results depended on the weights given to the HI and S (extended area) forecasts. This is evident in figure 5c which shows the RMSF scores for each merged run. Note that RMSF gives a direct measure of the rainfall rate accuracy where both the forecast and analysis have rain greater than a threshold value. All of the merged forecasts improve on the operational run, however the degree of improvement depends on the weights. It was discovered (table 5) that a weight of 1.0 for the HI forecast at T+15, 0.75 up to T+120, and those as for M6 up to T+360 would have given the best results. This is clearly evident in figure 5d.

3. Merging experiments

Operationally in Nimrod, extrapolation and UM precipitation forecasts are merged using a weighted combination, ie $R_{merged} = w_e R_{extrapolation} + w_m R_{model}$. The weights for the extrapolation are specified to decrease exponentially with lead time from 0.8 at T+1 to 0.2 by T+6. The weights for the model forecast are calculated according to the relative accuracy of the model rainfall forecast at T+0 and a one hour extrapolation from T-1. The score used to calculate the accuracy is based on cross-correlation weighted according to the size of each large object. This method of merging has two obvious limitations; the cross-correlation takes no account of rainfall rate accuracy, just location, and fixing the accuracy of the extrapolation at T+1 and T+6 may not always be appropriate as it is likely to vary according to the synoptic situation. In order to try and overcome these difficulties a new merging scheme based on the error variance of the extrapolation and UM components was devised and tested in parallel with the operational scheme.

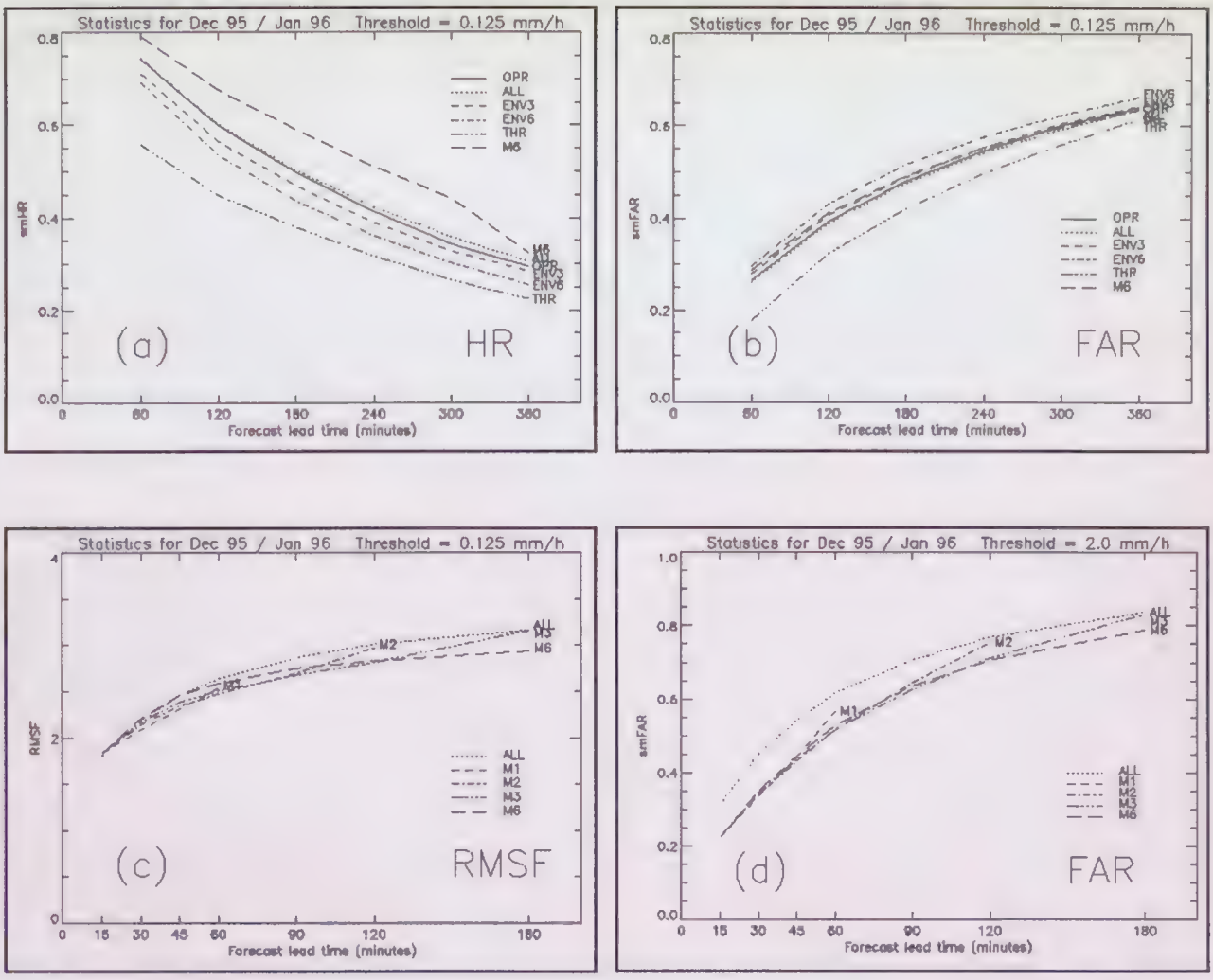

Figure 5. Graphs showing performance of experimental forecasts with lead time. (a) - Hit Rate threshold of 0.125 mm/h, (b) - False Alarm Rate, (c) Root Mean Square Factor, (d) - FAR threshold 2.0 mm/h. All experiments are labelled as explained in the text except ENV3 - 3 hour envelope and ENV6 - 6hour envelope.

3.1 Experimental merging scheme

The procedure adopted was as follows:

1. At T+0 identify the best extrapolation forecast method (envelope, normal, merged).

2. For the best extrapolation method calculate error variances (EV) according to forecast lead time using the union of the forecast and T+0 rainfall fields within the useable radar area at a threshold of 0.125 mm/h $(RMSF_OR - 1.0)^2$.

3. Determine the error variance for the latest available model rainfall forecast field verifying at T+0.

4. The weights for the merging are derived as follows:

$$W_{extrap}(t) = (EV_{model}) / (EV_{model} + EV_{extrap}(t))$$

$$W_{model}(t) = (EV_{extrap}(t)) / (EV_{model} + EV_{extrap}(t))$$

where t = 15 -> 360 minutes at 15 minute intervals.

5. The forecasts are then combined using the weights

$$R(t) = W_{extrap}(t) \times E(t) + W_{model}(t) \times M(t)$$

To identify the best extrapolation forecast method care had to be taken not to bias the outcome. The merged extrapolation generally performed extremely well at short range and envelope forecasts were better at longer ranges (> T+180). To avoid any complication of severely mis-matching forecasts it was decided to choose one extrapolation forecast type for any occasion rather than blending separate forecasts. To select the "best" forecast T+180 RMSF_OR errors were compared for the forecast for T+0. The threshold used was 0.125 mm/h. Three forecasts were available for selection: ALL - the operational object motion analysis over the entire Nimrod area, E3 - an envelope forecast using a time lag of 3 hours for object-matching, and MH - a modified merged extrapolation. MH was similar to M6 except that weights given to the high-intensity component (HI) were modified to give optimum results. At T+15 the HI weight was set to 1.0, 0.75 was used between T+30 and T+120 and then the same weights as for M6 thereafter.

In the experiment the error variance for the UM was kept constant at the value of the latest available forecast for T+0. The justification for this can be seen in table 6.

RMSF_OR errors in period 22/12/95 to 16/01/96 for lead times between T+60 and T+900 mins															
	60	120	180	240	300	360	420	480	540	600	660	720	780	840	900
M6	4.04	5.18	5.89	6.46	6.89	7.47	-	-	-	-	-	-	-	-	-
UM	-	-	6.26	6.36	6.40	6.42	6.48	6.46	6.39	6.48	6.63	6.98	6.85	6.83	6.69

Table 6. RMSF_OR errors for the same forecast data times used in extrapolation experiments between 22/12/95 and 16/01/96 for the UM and M6 forecasts at lead times between T+60 and T+900. Rainfall threshold used was 0.125 mm/h.

UM forecasts with lead times less than 180 minutes were not used in the experiments as the forecasts generally do not become available until after T+120. The UM errors vary little with lead time, exhibiting a small and erratic increase with increasing time. Conversely, M6 errors increase steadily. Admittedly by keeping the UM error variance constant there is a danger in using too much of the UM forecast, particularly when T+0 is close to the model data time, however, table 6 suggests that any adverse effects are likely to be small. An alternative approach would be to use the error variances appropriate to the UM lead time. However, this would mean calculating these from the previous model run at verification times 6 hours before T+0 for a UM cycle of 4 runs per day. This could produce very misleading weights if the synoptic situation changes significantly in the 6 hour period. However, it may be worth reconsidering when the number of UM runs is increased and if sufficient CPU and run time are available.

For the extrapolation forecasts, error variances were calculated at 15 minute intervals for lead times up to T+60 and then at T+90, T+120, and every 60 minutes to T+360. Intervening weights were derived by

linear interpolation. This was considered sufficient for the purposes of the experiment, where, beyond T+120 only hourly fields were scrutinised.

3.2 Results

The experiment was conducted using the same computer scripts as for the extrapolation experiments in section 2. Forecasts were run for periods verifying between 19/1/96 0600 UTC - 30/1/96 0600 UTC, 02/02/96 0900 UTC - 02/02/96 1700 UTC, 06/02/96 1200 UTC - 25/02/96 2300 UTC . Gaps in the period were due to problems arising after the NIMDEV development machine was switched temporarily to operational mode. Forecasts were produced using the technique described in section 3.1 (MHG2) and were compared with operational forecasts (OPM) and a forecast produced combining the UM with the MH extrapolation forecast method using the operational merging scheme (MHGE). Example forecasts from this trial period are shown in figures 6 and 7 below.

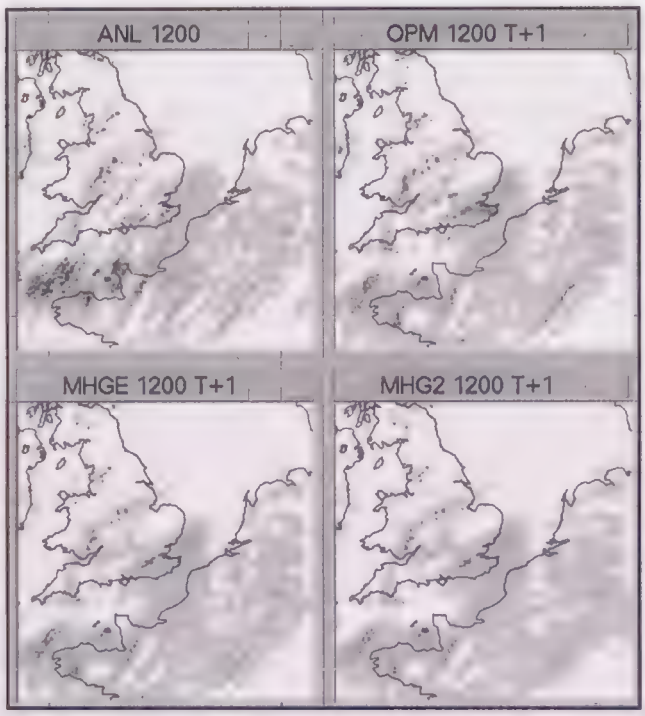

Figure 6. Analysis and T+1 forecasts of precipitation rate for 1200 UTC 20/2/96. "ANL" is the Nimrod analysis, "OPM" is the operational merged forecast, "MHGE" is the operational merging method used with the "MH" extrapolation forecast and "MHG2" is the experimental merging method used with the "best" extrapolation forecast. Light grey = 1/32 - 0.5 mm/h, mid-grey = 0.5 - 1.0 mm/h and dark grey > 1.0 mm/h.

Figure 6 shows analyses and T+1 forecasts for 1200 GMT 20/2/96. The synoptic situation was a very cold northeasterly airflow with areas of snow moving quickly southwestwards. Over the UK the main error in the operational forecast was that there was too much moderate snow predicted in a line from Kent to London to Surrey. A less significant error was that precipitation was pushed too far west into Wales. Both of the experimental forecasts improved upon the operational prediction. The heavier precipitation over Kent and

Surrey is more correctly aligned NE/SW and the handling of the heavier precipitation in Wales is also better. Other differences are very small and there are only minor differences between the two experimental merged forecasts. The best extrapolation method in this case was the MH forecast and the weight given to this forecast in the MHG2 method was 0.86, which was higher than the operational value. The improvement can be attributed almost entirely to the MH forecast being better than the operational extrapolation. Moreover it is the movement of the heavier precipitation which is better. However, the T+3 forecast for 1200 UTC shown in figure 7 illustrates a different effect.

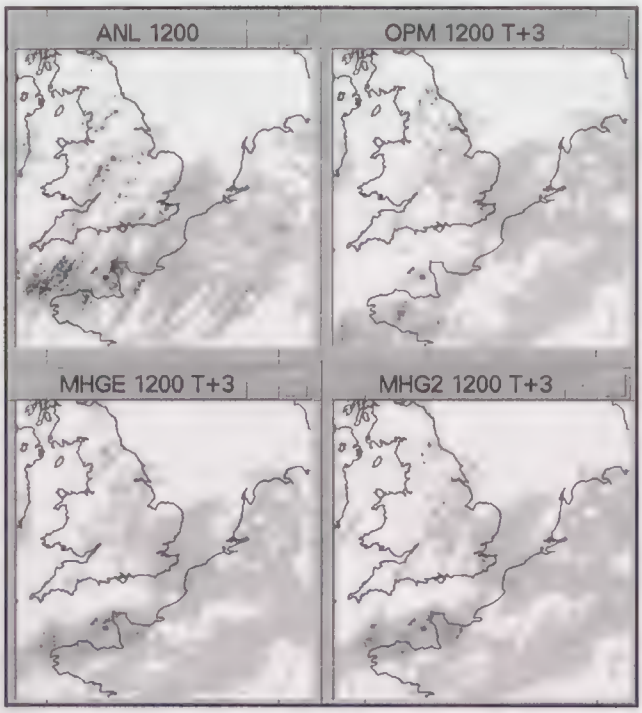

Figure 7. Analysis and T+3 forecasts of precipitation rate for 1200 UTC 20/2/96. "ANL" is the Nimrod analysis, "OPM" is the operational merged forecast, "MHGE" is the operational merging method used with the "MH" extrapolation forecast and "MHG2" is the experimental merging method used with the "best" extrapolation forecast. Light grey = 1/32 - 0.5 mm/h, mid-grey = 0.5 - 1.0 mm/h and dark grey > 1.0 mm/h.

At this time there is a major error in the operational forecast over NW France and the Channel Islands. The forecast is too fast whereas over the UK the differences are smaller. The experimental forecast MHG2 chose the three hour envelope method (E3) as the best forecast method and the weight given to the extrapolation was 0.71 which was higher than in the operational forecast. The reason why the MHG2 forecast was much better than OPM was that the snow was basically convective in origin and was continually being formed on the northern edge after sufficient fetch across the English Channel in the strong northeasterly flow, which is exactly the sort of situation the envelope technique was designed for. The MHGE forecast was slightly faster than MHG2 but still produced a better forecast than OPM as at 0900 the rainfall object was not entirely within the radar area, consequently the object motion analysis over the whole area was able to give a more appropriate velocity which turned out to be close to the envelope values.

3.2.1 Objective verification

Of all the merged MHG2 forecasts made in the test period, 85% used the MH method as the extrapolation component, 12% E3 (3 hour envelope) and 3% the operational method over the extended area (A). On the 20th February the E3 method was chosen for 9 out of 24 forecasts. The experimental average and maximum weights, and the operational weights for the extrapolation component in the period as a function of lead time were:

Lead time (mins)	Average	Maximum	Operational
T+15	0.82	1.00	0.98
T+30	0.74	0.96	0.92
T+45	0.67	0.94	0.86
T+60	0.62	0.92	0.80
T+90	0.56	0.96	0.70
T+120	0.52	0.92	0.61
T+180	0.47	0.94	0.46
T+240	0.42	0.95	0.35
T+300	0.39	0.91	0.26
T+360	0.35	0.90	0.20

These figures indicate that weights given to the extrapolation component cover a smaller range between T+15 and T+360 than those used operationally. On average more weight is given to the UM up to T+180 in the MHG2 forecast than operationally and more weight to the extrapolation in the later stages. The maximum weight values show that on occasions the UM is very poor and most weight is then given to the extrapolation component. The fact that the basic extrapolation forecast (ALL) was selected as the best on only 3% of occasions provides a strong justification for running the MH forecast. The inclusion of an envelope based extrapolation is also justified.

The merged forecasts were verified objectively in the same way as the extrapolations previously. To provide a control, all the operational merged forecast statistics were assembled for the same forecast run times used in the experiment.

Over the period, forecast MH showed a marginal (less than 3%) improvement over M6 between T+30 and T+120, particularly for thresholds above 2.0 mm/h. The envelope forecasts exhibited the same characteristics as in the previous period, performing on average relatively better after T+180 than before T+180. Verification statistics for HR, FAR and RMSF_OR for threshold ranges of 0.125 and 2.0 mm/h in forecasts A, MH, OPM, MHGE and MHG2 are shown in tables 7 and 8 below.

The scores for the A forecast were similar to and slightly worse than those for the previous experiment. This probably suggests that the synoptic situations in this period were more challenging. Comparing the extrapolation forecasts only (A, MH) we also come to the same conclusions as in the previous experiment. That is, for rain above 2.0 mm/h the MH forecast has a much reduced FAR but a slightly worse HR and the converse for rain above 0.125 mm/h.

For rain above 0.125 mm/h threshold, the RMSF_OR scores show that forecast MHG2 performed best after T+30, before then MHGE had the lowest score. However, close examination of the HR and FARs shows that for HR, MHG2 is better than MHGE after T+30, but for FAR MHG2 is substantially worse at and before T+120. This is consistent with subjective assessments which noted a tendency for too much lighter rain

Table 7). HR, FAR and RMSF_OR for the full area (A), MH, OPM, MHGE, and MHG2 forecasts for a threshold rainfall rate of 0.125 mm/h. HR and FAR are expressed as percentages. Shading indicates where the HR and FAR meet the mandatory Nimrod User Requirement.

Threshold = 0.125 mm/h															
Lead time (min)	Hit Rate (HR)					False Alarm Rate (FAR)					RMSF_OR (x100)				
	A	MH	OPM	MH GE	MH G2	A	MH	OPM	MH GE	MH G2	A	MH	OPM	MH GE	MH G2
T+360	29	30	53	56	57	62	62	60	60	61	687	701	577	567	565
T+300	33	39	53	57	59	59	60	57	58	59	702	696	581	559	551
T+240	38	44	53	59	62	55	56	55	56	58	711	672	589	551	534
T+180	45	51	54	61	65	49	52	51	52	55	700	618	592	539	508
T+120	55	61	58	65	70	41	45	44	45	51	632	546	567	507	474
T+90	61	67	-	70	74	36	39	-	40	48	572	495	-	472	443
T+60	70	75	69	76	79	28	32	31	32	43	494	427	474	417	399
T+45	76	80	-	81	82	24	27	-	28	38	439	379	-	374	366
T+30	82	85	-	85	85	18	21	-	21	31	368	319	-	316	321
T+15	89	90	-	91	89	12	14	-	14	21	273	252	-	251	263

Table 8). HR, FAR and RMSF_OR for the full area (A), MH, MHGE, and MHG2 forecasts for threshold rainfall rate of 2.0 mm/h. HR and FAR are expressed as percentages.

Threshold = 2.0 mm/h												
Lead time (min)	Hit Rate (HR)				False Alarm Rate (FAR)				RMSF_OR (x100)			
	ALL	MH	MH GE	MH G2	ALL	MH	MH GE	MH G2	ALL	MH	MH GE	MH G2
T+360	9	9	7	11	87	87	69	73	313	313	297	281
T+300	11	10	8	12	85	82	66	69	306	303	293	278
T+240	13	11	9	12	83	77	64	68	301	296	289	276
T+180	18	15	12	13	78	70	59	63	291	286	283	276
T+120	25	22	17	16	72	64	59	58	281	275	271	270
T+90	31	27	23	18	67	59	55	53	271	263	261	264
T+60	42	38	34	25	58	48	46	44	254	243	242	249
T+45	48	44	41	30	52	42	40	39	239	227	226	235
T+30	59	55	53	41	44	33	32	31	219	206	205	216
T+15	70	70	70	58	32	24	23	23	185	176	175	185

in the MHG2 forecasts at T+60. The reason is undoubtedly the greater influence of the UM in the MHG2 forecast than in MHGE. Beyond T+180 the relatively greater influence of the extrapolation in MHG2 reduces the FAR difference. By comparing the MH and MHGE figures we can see that for the 0.125 mm/h threshold the benefit of the UM is to increase the HR after T+60. However, for the heavier rain, > 2.0 mm/h the benefit is in decreasing the FAR. Up to T+90 the higher weight given on average to the extrapolation in the MHGE forecast leads to a much improved forecast in terms of HR than the MHG2 forecast. This is also borne out in the RMSF_OR scores in table 8 which show the MHGE forecast as best up to T+90 and MHG2 better beyond T+90.

Comparing these results with those from the operational merged forecast in table 7 over the same sample we see that both of the experimental merged forecasts improve on the operational forecast for HR for all lead times especially before T+180. It is also noticeable that had the experimental forecasts been used operationally we would have met the mandatory UR for HR at all times. For FAR the differences are small with the operational scheme slightly better. For RMSF_OR the experimental forecasts show a substantial improvement over the operational system. It is also interesting to note that the scores for for the MH extrapolation forecast are better than OPM up to and including T+120. The main difference is in HR with the FARs very similar.

4. Conclusions and recommendations

On the basis of the evidence presented in this report and of routine examination of outputs from the experimental forecasts (available as hard copy) the following conclusions and recommendations are made.

1. The difference between the operational extrapolation forecasts and those made from an extended object analysis are negligible with the latter slightly better. It has also been demonstrated that extending the area of motion analysis can provide noticeable improvement in forecast output. It is therefore recommended that the object motion analysis be performed over the entire Nimrod area and that this be implemented as soon as possible. Changes to the code will be small and the maximum permitted number of large objects should be increased from the current value of five to the highest number possible on the Nimrod computer (probably seven).

2. The method of extrapolating precipitation according to rate and then merging the forecasts to provide a combined extrapolation (section 2.4) should also be implemented as soon as possible. The best way of combining the forecasts is to use the weights used for the MH forecast. The expected benefit can be a 12% improvement in extrapolation accuracy for rain above 0.125 mm/h up to T+180, and a 4% improvement for rain above 2.0 mm/h as measured using RMSF_OR. The incidence of false alarms for rain above 2.0 mm/h will be significantly reduced at all stages of the forecasts.

3. The operational method of merging the extrapolation with the UM should be retained for the time being. Merging the UM with the MH forecast provides improved forecasts at all lead times. However, the method of deciding the best extrapolation forecast should be incorporated and an envelope extrapolation should be routinely produced. The benefit of using envelope velocities will be seen on a case by case basis, especially in showery situations where showers are persistent and restricted to geographical regions. In not opting for implementation of the experimental merging technique at this stage, some potential benefit at T+180 and beyond will be lost, however, this must be balanced against the probability that the shorter range forecasts would become worse, particularly for heavier rain.

4. The results obtained so far on examining different ways of merging extrapolation and UM forecasts has been encouraging. However, the problem remains that not enough weight is given to the extrapolation in the early stages of the forecast, particulary now that the extrapolation has been improved. Further work needs to be done to refine the method for determining weights, which should be done in conjunction with work already underway to improve the UM presentation of precipitation on the Nimrod grid. Subsequent

inclusion of an improved model merging method is still recommended subject to improvements being demonstrated in the shorter range nowcasts.

To implement recommendations 1 - 3 would not require major programming changes. However, two additional object motion analyses and forecasts would have to be run and saved, (high intensity and envelope). The inclusion of these extra forecasts can be done with minor modifications to existing programs. The merging of the high intensity and standard extrapolation forecasts would be a new routine, which has already been written to Nimrod standards. Deciding on the best extrapolation to merge with the model would require minor modifications to the existing merging program.

5. References

Conway, B.J., G Ryall, N Thomas, "Summary Report on tests of the automated rainfall extrapolation/advection scheme for Nimrod".
Nowcasting Development Group Technical Note 64.